AF/ZJW

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of

Hyun-doo SHIN, et al.

Appln. No.: 09/823,272

Confirmation No.: 7285

Filed: April 2, 2001

For: INDEXING METHOD OF FEATURE VECTOR DATA SPACE

Docket No: Q59549

Group Art Unit: 2625

Examiner: YUBIN HUNG

SUBMISSION OF APPEAL BRIEF

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith please find an Appeal Brief. A check for the statutory fee of $500.00 is attached. The USPTO is directed and authorized to charge all required fees, except for the Issue Fee and the Publication Fee, to Deposit Account No. 19-4880. Please also credit any overpayments to said Deposit Account. A duplicate copy of this paper is attached.

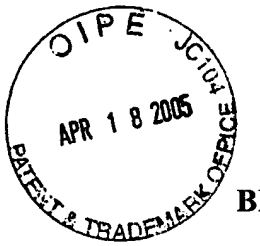
Respectfully submitted,

Chid S. Iyer
Registration No. 43,355

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON OFFICE
23373
CUSTOMER NUMBER

Date: April 18, 2005



PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of

Hyun-doo SHIN, et al.

Docket No: Q59549

Appln. No.: 09/823,272

Group Art Unit: 2625

Confirmation No.: 7285

Examiner: YUBIN HUNG

Filed: April 2, 2001

For: INDEXING METHOD OF FEATURE VECTOR DATA SPACE

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 41.37 Appellants are submitting an Appeal Brief to appeal from the Final Office Action dated October 15, 2004 (hereinafter "the Final Office Action"), wherein claims 1-13 are finally rejected. This Appeal Brief is accompanied by a Submission which includes the required appeal fee set forth in 37 C.F.R. § 41.20(b)(2). Appellants' Notice of Appeal was filed on January 18, 2005. Therefore, the present Appeal Brief filed is accompanied with a one-month extension of time fee.

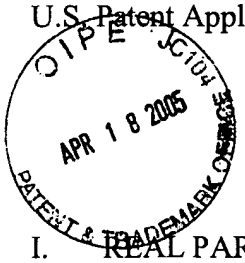


Table of Contents

I. REAL PARTY IN INTEREST

The real parties in interest are SAMSUNG ELECTRONICS CO., LTD. and THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (Assignees) by virtue of an assignment executed by Mr. Hyun-Doo SHIN and Mr. Yang-Lim CHOI, the co-inventors on July 4, 2001 (SAMSUNG ELECTRONICS), and recorded by the Assignment Branch of the U.S. Patent and Trademark Office on July 9, 2001 (at Reel 011961, Frame 0824) and by virtue of an assignment executed by Mr. Bangalore S. MANJUNATH and Mr.Peng WU, the co-inventors on June 21, 2001 and June 27, 2001 (REGENTS OF THE UNIVERSITY OF CALIFORNIA), and recorded by the Assignment Branch of the U.S. Patent and Trademark Office on July 9, 2001 (at Reel 011961, Frame 0828).

II. RELATED APPEALS AND INTERFERENCES

Upon information and belief, there are no other prior or pending appeals, interferences, or judicial proceedings known to Appellants, Appellants' representatives or the Assignees that may be related to, be directly affected by, or have a bearing on the Board's decision in this appeal.

III. STATUS OF CLAIMS

Each of pending claims 1-13 is finally rejected (*see* Final Office Action dated October 15, 2004). By a section 116 amendment filed on April 15, 2005, claims 2 and 9 are cancelled. Consequently, claims 1, 3-8, 10-13 are the claims on appeal (*see* Claims Appendix).

IV. STATUS OF AMENDMENTS

One Final Office Action (dated October 15, 2004) is of record in the current appeal. A 116 amendment was filed on April 15, 2005 with clarifying amendments.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

Background

The present invention relates to an indexing method of a feature vector data space, and more particularly, to an indexing method for finely indexing cells having a high concentration of feature vectors by hierarchically approximating the feature vectors depending on their distribution in a feature vector data space.

Fast and efficient access to a database is desired for a significant number of applications commonly used today. Specifically, managing such databases that include multimedia data and providing methods to access the multimedia data has become increasingly important. In such multimedia data, a few hundred thousand to a few million or more items is not uncommon. For each object (or record) in such a multimedia database, the degree of the data (for example, the dimensionality of the attributes in the data) is much higher than that in a conventional database.

Feature vectors are a convenient way of representing such high degree, multi-dimensional data. A collection of feature vectors that form part of an application can be termed as a feature vector data space.

Efficient indexing methods are required to access a database that stores feature vectors. Some indexing methods aim at minimizing storage overhead, while others may focus on efficiently supporting the range of queries. The present invention is aimed at provide an indexing method for a database that stores a feature vector data space. Further, in the present invention, the feature vectors are classified into uniform-sized cells and cells in which a high concentration of feature vectors exist are more finely indexed.

A computer-readable recording medium for storing program codes used for performing the indexing method of a feature vector data space as well as a similarity searching method of performing a similarity search in such a feature vector data space are also provided for the present invention.

Claims 1, 3-6

Claims 1, 3-6 are directed to indexing method of a feature vector data space in which a plurality of feature vectors are indexed. The claims reciting *inter alia*: partitioning the feature vector data space into a plurality of cells having a uniform size, determining whether one or more cells, on each of which one or more of said plurality of feature vectors are correspondingly concentrated, exist; and hierarchically indexing the feature vector data space when it is determined that said one or more cells, on each of which said one or more of said plurality of feature vectors are correspondingly concentrated, exist. One or more feature vectors are said to be concentrated in a cell when the cell contains more feature vectors than a predetermiend threshold.

Claims 7, 8, 10 and 11

Claims 7, 8, 10 and11 recite a computer programming product comprising instructions to implement the method discussed above.

Claims 12-13

Claims 12-13 recite methods for searching for similarity in a feature vector data space in which feature vectors are indexed. The claims recite *inter alia:* performing a similarity search in the feature vector data space, which has been indexed, by determining whether each of one or more cells, on which the feature vectors are correspondingly concentrated, exists and

hierarchically indexing the feature vectors in said each of one or more cells, on which it is determined that the feature vectors are correspondingly concentrated, according to a predetermined indexing method.

Means-Plus-Function Claims

No means-plus-function or step-plus-function have been identified among the claims on appeal.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

1. Claims 1, 7, 12 and 13 are rejected under 35 U.S.C. § 102(e) as allegedly being anticipated by Kothuri et al. (US 6,381,605).

2. Claims 2 and 9 are rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Kothuri et al. (US 6,381,605) as applied to claims 1, 7, 12 and 13 above, and further in view of "2n-Tree Classifiers" in IBM Technical Disclosure Bulletin, Vol. 34, No. 4B, September 1991, pp. 225-228 [hereinafter "the IBM reference"].

3. Claim 3 is rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Kothuri et al. (US 6,381,605) as applied to claims 1, 7, 12 and 13 above, and further in view of Massen (US 5,809,165).

4. Claims 4-6, 8, 10 and 12-13 are rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Kothuri et al. (US 6,381,605) as applied to claims 1, 7, 12 and 13 above, and further in view of "A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces," Proceedings of the 24th International Conference on Very Large Data Base, New York, August 1998, pp. 194-205 [hereinafter "the ICVLDB reference"].

5. Claim 11 is rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Kothuri et al. (US 6,381,605) and Weber et al. as applied to claims 4-6, 8, and 10 above, and further in view of Massen (US 5,809,165).

VII. ARGUMENT

1. Claims 1, 7, 12 and 13 are not anticipated by Kothuri

The present invention, as recited in claim 1, requires partitioning the feature vector data space into a plurality of cells having a uniform size. Further, a test is performed to determine if cells where one or more feature vectors are concentrated exist. If so, the hierarchical indexing is performed on the feature vector data space. One or more feature vectors are considered to be concentrated in a cell when the cell contains more feture vectors than a predetermiend threshold

To clarify further, a 116 amendment with clarifying amendments to claims 1, 7 and 12 was filed on April 15, 2005.

In rejecting claim 1, the Examiner refers to Figs. 3 & 5, numerals 506 and 518 and 14-55-15:43 of Kothuri in alleged support for determining whether one or more cells, on each of which one of more of said plurality of feature vectors are correspondingly concentrated exist.

Further the Examiner refers to 3:27-37 and 14:55-56 in alleged support for hierarchically indexing the feature vector data space if such one or more cells exist.

As an initial matter, Kothuri does not disclose any feature vector space. Kothuri is related to indexing of multi-dimensional or multi-attribute data. While feature vectors could have more than one dimensions, a general teaching on multi-dimensional data cannot be considered to be a specific disclosure related to a feature vector space where the number of dimensions are significantly higher.

Kothuri follows a completely different approach for hierarchically indexing multi-dimensional data. Even the individual steps in the hierarchical indexing do not refer to determining cells in which feature vectors are concentrated.

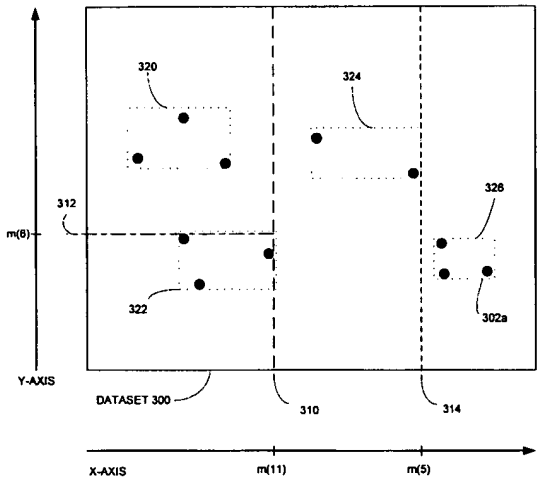

320
324
312
326
m(6)
322
302a
Y-AXIS
DATASET 300
310
314
X-AXIS
m(11)
m(5)

FIG. 3

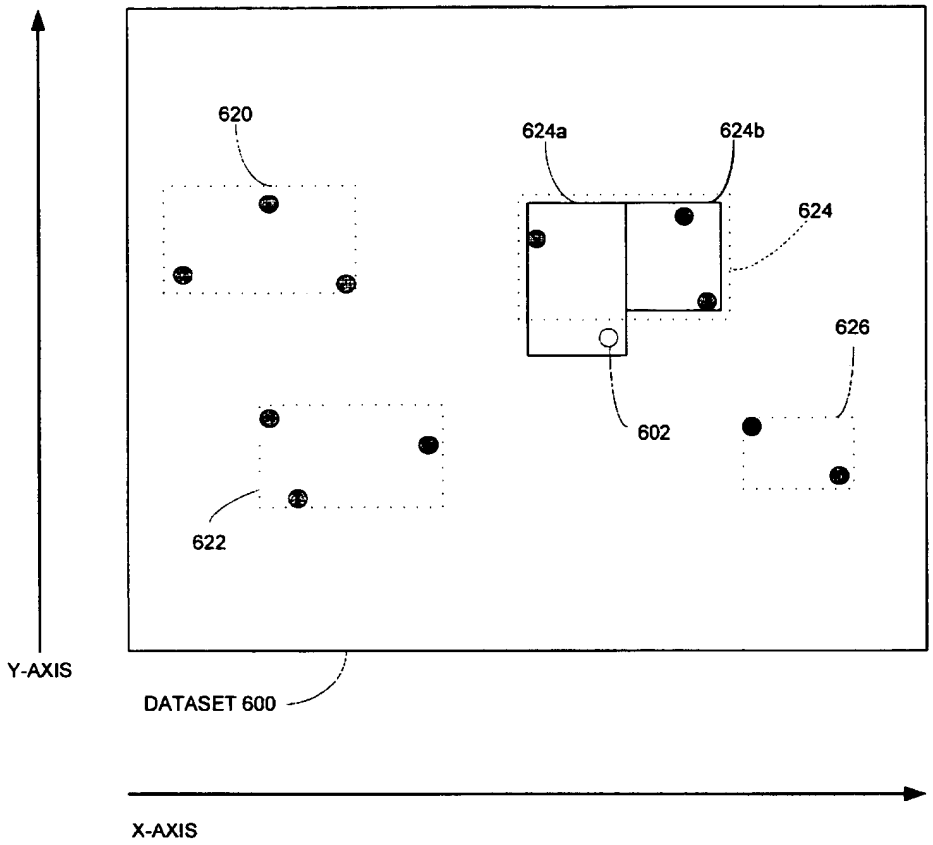


FIG. 6A

Referring to Figs 3 and 5 of Kothuri, dataset 300 comprises multiple data points that vary more in the x dimension than the y dimension. They are sorted according to their values in the x dimension. An effective location for a first division is determined based on a median computation, which is illustrated by dividing line 310. After the first partition, it is determined whether each new subset of data will fit into a single R-tree node. Each subset that is too large

(i.e., each subset that contains more than M data points) is further divided in a similar manner. Further divisions are performed till each subset can fit into a single node.

As described in the flowchart of Fig. 5, in step 506 it is determined if the data items fit into one node. If they do not fit, the variance in each dimension is determined. Then a dimension or attribute hierarchy having a greatest variance is selected and the data items sorted in that dimension. In step 514 the data items are divided in the selected dimension into two or more subsets. An approximate median may be computed as described above in order to divide the data items in half as nearly as possible. Alternatively, the data items may be divided into a number of clusters, each of which contains a number of data items that will fit into one leaf node.

As noted above, clearly, the emphasis is on dividing the data into subsets such that each data will fit into a leaf node. On the other hand, in the present invention, as recited in claim 1, as amended, the cells are divided into a plurality of cells having a **uniform size** regardless of whether they will fit into a particular cell or sub-cell. After that it is determined if cells where a plurality of feature vector space are concentrated exist. This is done by determining if there are more number of cells than a predetermined threshold. If they exist, then the vector data space is hierarchically indexed.

The difference between the present invention and Kothuri are clear from the respective illustrations in Fig. 3 and 6A of Kothuri and Figs 2-3 of the present Specification.

As shown in Fig. 3, based on variance information, Kothuri decides how the data is subdivided. For example, lines 310, 312 and 314 are drawn at appropriate locations based on the variance of the data. Then sub-divisions that have more data are then subdivided further, again based on variance. In Fig. 6A, clusters 620, 622, 624 and 626 of data are generated based on the amount of data that can be stored in a leaf node.

On the other hand, as shown in Fig. 2 of the present Specification, the data is divided into uniform cells 00 00, 00 01, etc. Further, the cells 20 and 22 that have feature vectors concentrated are determined. Then these cells are further divided as shown in Figs 3A and 3B.

Even if a node of Kothuri is read on the cell/sub-cell of the present invention, as the Examiner appears to be doing, the above steps of Kothuri are significantly different from determining the cells in which feature vectors are concentrated.

To anticipate a claim, each and every element and its limitations must be disclosed by the cited reference. Claim 1 should be allowed at least because Kothuri does not disclose (or suggest) dividing the data into uniform cells, performing hierarchical indexing if it is determined that one or more cells with one or more feature vectors concentrated exist. In fact, the purpose of the present invention is to provide a finer indexing for those feature vectors that get concentrated in certain cells. If the data items are simply divided into clusters or divided so as to fit into leaf nodes, as in Kothuri, such finer indexing can not be provided.

Claims 7, 12 and 13 include limitations analogous to the ones discussed above in relation to claim 1.

Rejection of claims 2 and 9 based on the IBM reference "2n-Tree Classifiers"

The above rejection is rendered moot since these claims are canceled. However, the limitations (or analogous limitations) in these claims are added to claims 1, 7 and 12. Therefore, the differences between these added limitations and the IBM reference are discussed herein. The Examiner incorrectly alleges that the IBM reference suggests dividing the feature vector space into classes of uniform size. In fact, the Examiner is believed to be mischaracterizing the teachings of the IBM reference.

The IBM reference suggests a tree of hypercubes. If the feature vector is a m bit integer, then the hypercube has a side length of 2m. The root node corresponds to the entire feature vector space. Its children are equal-sized hypercubes with size 2m-1, each of which again have children with sizes 2m-2 and so on. During classification the first level of the tree is based on the highest order of the bits in the m-bit integer. The second level of the tree is based on the next order bit and so on. This is followed recursively till the leaf is reached where the particular feature vector is located. More specifically, the feature vector in IBM is an m-bit integer, with m being the number of attributes. Further, the size of the hypercube at various level of the tree are 2m, 2m-1, 2m-2, etc. Therefore at each level of the tree, a particular attribute is distributed across all the branches at that level. Further, the size is based on the number of attributes. On the other hand, a division of feature vectors into uniform sized classes is based on the value range of the attributes as opposed to the number of attributes.

As can be seen, the reference to equal sized hypercubes in the context of the IBM reference is very different from dividing the data into uniform sized classes as in the present invention. As best understood, in the IBM reference, the classification process is such that the leaf node will always have only one feature vector. Moreover, in the classification scheme followed by IBM, a particular path in the tree, if traversed, will lead to only one feature vector. On the other hand, there is no division based on uniform class sizes as in the present invention.

A skilled artisan would not have found it obvious to combine the teachings of Kothuri and the IBM reference. Significantly, even if the teachings were combined, the deficiency noted above in the teaching of Kothuri will not be overcome.

Claim 3 and 11 are not obvious over Kothuri and Massen

Claim 3 is dependant on claim 1, and should be allowed at least for the same reasons. Claim 11 is dependant on claim 7, and should be allowed at least for the same reasons. Further, Massen does not overcome the deficiency noted above in the teachings of Kothuri and the IBM reference.

Claims 4-6, 8, 10 and 12-13 are not obvious over Kothuri and the ICVLDB reference

Claims 4-6 are dependant on claim 1, and should be allowed at least for the same reasons. Likewise claims 8 and 10 are dependant on claim 7. Further, the ICVLDB reference does not overcome the deficiency noted above in the teachings of Kothuri and the IBM reference.

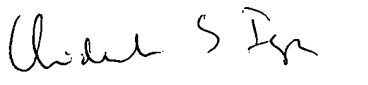
CONCLUSION

For the above reasons as well as the reasons set forth in Appeal Brief, Appellant respectfully requests that the Board reverse the Examiner's rejections of all claims on Appeal. An early and favorable decision on the merits of this Appeal is respectfully requested.

Unless a check is submitted herewith for the fee required under 37 C.F.R. §41.37(a) and 1.17(c), please charge said fee to Deposit Account No. 19-4880.

The USPTO is directed and authorized to charge all required fees, except for the Issue Fee and the Publication Fee, to Deposit Account No. 19-4880. Please also credit any overpayments to said Deposit Account.

Respectfully submitted,



Chid S. Iyer
Registration No. 43,355

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON OFFICE
23373
CUSTOMER NUMBER

Date: April 18, 2005

CLAIMS APPENDIX

CLAIMS 1, 3-8, 10-13 ON APPEAL:

1. An indexing method of a feature vector data space in which a plurality of feature vectors are indexed, the indexing method comprising the steps of:

(p-a1) partitioning the feature vector data space into a plurality of cells having a uniform size;

(a) determining whether one or more cells, from said plurality of cells, on each of which one or more of said plurality of feature vectors are correspondingly concentrated, exist; and

(b) hierarchically indexing the feature vector data space when it is determined that said one or more cells, on each of which said one or more of said plurality of feature vectors are correspondingly concentrated, exist in the step (a)

wherein, one or more feature vectors are concetrated in a cell when the cell contains more feature vectors than a predetermined thresold.

3. (original): The indexing method of claim 1, wherein the step (a) comprises the sub-steps of:

(a-1) constructing a histogram illustrating a number of said plurality of feature vectors in each of a plurality of cells, including said one or more cells; and

(a-2) analyzing a distribution of said plurality of feature vectors using the histogram and determining whether said one or more cells, on each of which said one or more of said plurality of feature vectors are correspondingly concentrated, exist.

4. (original): The indexing method of claim 1, wherein the step (b) comprises the step of indexing the feature vector data space using a vector approximation file.

5. (original): The indexing method of claim 4, wherein the step (b) comprises the sub-steps of:

(b-1) constructing a sub-vector approximation file over each of said one or more cells, on which said one or more of said plurality of feature vectors are correspondingly concentrated; and

(b-2) approximating said one or more of said plurality of feature vectors in said each of said one or more cells, on which said one or more of said plurality of feature vectors are correspondingly concentrated, using the vector approximation file and a corresponding sub-vector approximation file.

6. (original): The indexing method of claim 1, wherein the step (b) comprises the sub-steps of:

(b-1) partitioning each of said one or more cells into a corresponding plurality of sub-cells, when it is determined that said each of said one or more cells, on which said one or more of said plurality of feature vectors are correspondingly concentrated, exists in the step (a); and

(b-2) approximating said one or more of said plurality of feature vectors in said each of said one or more cells, using said corresponding plurality of sub-cells, thereby hierarchically indexing the feature vector data space.

7. (original): A computer-readable recording medium for storing program codes for performing an indexing method of a feature vector data space in which a plurality of feature vectors are indexed, the indexing method comprising the steps of:

(p-a1) partitioning the feature vector data space into a plurality of cells having a uniform size;

(a) determining whether one or more cells, from said plurality of cells, on each of which one or more of said plurality of feature vectors are correspondingly concentrated, exist; and

(b) hierarchically indexing the feature vector data space when it is determined that said one or more cells, on each of which said one or more of said plurality of feature vectors are correspondingly concentrated, exist in the step (a)

wherein, one or more feature vectors are concetrated in a cell when the cell contains more feature vectors than a predetermined thresold.

8. (original): The computer-readable recording medium of claim 7, wherein the step (b) comprises the step of indexing the feature vector data space using a vector approximation file.

9. (canceled).

10. (original): The computer-readable recording medium of claim 7, wherein the step (b) comprises the sub-steps of:

(b-1) partitioning each of said one or more cells into a corresponding plurality of sub-cells, when it is determined that said each of said one or more cells, on which said one or more of said plurality of feature vectors are correspondingly concentrated, exists in the step (a); and

(b-2) approximating said one or more of said plurality of feature vectors in said each of said one or more cells, using said corresponding plurality of sub-cells, thereby hierarchically indexing the feature vector data space.

11. (original): The computer-readable recording medium of claim 8, wherein the step (a) comprises the sub-steps of:

(a-1) constructing a histogram illustrating a number of said plurality of feature vectors in each of a plurality of cells, including said one or more cells; and

(a-2) analyzing a distribution of said plurality of feature vectors using the histogram and determining whether said one or more cells, on each of which said one or more of said plurality of feature vectors are correspondingly concentrated, exist, and

the step (b) comprises the sub-steps of:

(b-1) constructing a sub-vector approximation file over each of said one or more cells, on which said one or more of said plurality of feature vectors are correspondingly concentrated; and

(b-2) approximating said one or more of said plurality of feature vectors in said each of said one or more cells, on which said one or more of said plurality of feature vectors are correspondingly concentrated, using the vector approximation file and a corresponding sub-vector approximation file.

12. (original): A method of searching for similarity in a feature vector data space in which feature vectors are indexed, the method comprising the step of (a) performing a similarity search in the feature vector data space, which has been indexed, by determining whether each of one or more cells, from among a plurality of uniformly sized cells, on which the feature vectors

are correspondingly concentrated, exists and hierarchically indexing the feature vectors in said each of one or more cells, on which it is determined that the feature vectors are correspondingly concentrated, according to a predetermined indexing method.

13. (original): The method of claim 12, wherein the step (a) is performed based on a nearest neighbor search.